

# Honorable Mention

## WHAT PUBLIC LIBRARY NATIONAL RATINGS SAY

RAY LYONS is an independent consultant living in Cleveland; raylyons@gmail.com. Ray is reading *Carter Beats the Devil* by Glen David Gold.

NEAL KASKE is Chief, Public Services and Regional Libraries Branch, The National Oceanic and Atmospheric Administration Central Library, Silver Spring, Maryland; nkaske@gmail.com. Neal is reading *Common Wealth: Economics for a Crowded Planet* by Jeffrey D. Sachs.

Nearly every year since 1999, one hundred public libraries have been rated "best in class" by Hennen's American Public Library Ratings (HAPLR).<sup>1</sup> Many of the libraries have celebrated the occasion with glowing press releases, in-house banners, and communiqués with boards of trustees, funding authorities, and citizens. It seems fortunate that HAPLR has been available to certify the excellence of these select libraries. With the *LJ* Index, a new rating system from *Library Journal*, on the horizon, libraries will have another opportunity to ponder what ratings might mean for them.

When libraries consult ratings systems such as HAPLR or the *LJ* Index, they might assume that the scores are accurate and conclusive. Perhaps they envision them as *quality points* precisely calculated for straightforward comparisons between libraries. Thomas Hennen has promoted his ratings this way, describing them as "composite average[s] that can be compared to the score[s] of all other public libraries," and "like a Scholastic Aptitude Test with a theoretical score between 1 and 1,000."<sup>2</sup> Measures involving *composite averages* and *theoretical scores* certainly seem like they would be statistically sound and trustworthy. By comparison, *LJ* Index takes an approach calling for more cautious interpretation of ratings and wider acknowledgement of their limitations.

How can these simple and inexpensive rating systems produce such definitive information about public library performance? How do the calculations work? How accurate are they? Do they measure library excellence and greatness? If so, how do they assess dimensions such as excellent customer service, well-trained staffs, responsiveness to community needs, quality and availability of materials, and others that have been espoused by Holt and by Childers and Van House?<sup>3</sup>

A study addressing these questions was recently conducted as part of a library and information science graduate internship that the primary author (Lyons) completed at the U.S. National Commission on Libraries and Information Science (NCLIS).<sup>4</sup> The coauthor (Kaske), who was director of surveys and statistics at the commis-

sion at the time, recommended that Lyons study the statistical foundations of HAPLR. The purpose here is to summarize the findings from the study. Some of the issues explored pertain to library ratings in general, including HAPLR and other approaches. These instances will be noted. Based on the study results, it is recommended that libraries reconsider their use of HAPLR as a measure of library performance, quality, greatness, or excellence. Several methodological shortcomings make HAPLR insufficient for these purposes:

- HAPLR is not a scientifically validated measurement instrument. Without validation testing, there is no basis for concluding that HAPLR assesses library quality, excellence, greatness, or value.
- HAPLR uses a set of library input/output measures that is too narrow for fully measuring library performance. As seen in table 1, HAPLR uses only five input and three output statistics. The ratings disregard service outcomes, library mission, service responses, alignment between programming and community needs, collection quality, and other key operational areas.
- Based on HAPLR calculations, 50 percent of a library's score is dependent on library input statistics.<sup>5</sup> It is generally agreed within the library profession that input statistics are inadequate measures of library performance, service excellence, programmatic relevance, quality, or value.
- The federal data that HAPLR uses are imprecise due to inconsistencies in local compilation of

the data, sampling error, imputation, and other sources. Because of inevitable "noise" in the data, any comparative ratings based on these statistics will be approximate at best.

- Even assuming HAPLR to be a valid and accurate measure of library excellence or greatness, the scores cannot indicate how much excellence or greatness libraries might possess. As a result, comparing different libraries' HAPLR scores in a single year, or the same library's scores in different years, is somewhat pointless. This problem will be discussed later in the article.

## Not Like SAT Scores

While Hennen does not claim that HAPLR is exactly like the SAT testing regimen, his association of the two measurement systems implies that they share certain characteristics in common. Upon closer examination, however, essential differences between HAPLR and SAT become evident. First of all, the College Entrance Examination Board (CEEB), creator of the SAT examinations, has conducted numerous scientifically designed studies to assess the validity of the exams. By validity we mean what the field of behavioral science research calls measurement validity, which has been defined as: "the extent to which any measuring instrument measures what it is intended to measure. . . . One validates not the measuring instrument itself but the instrument in relation to the purpose for which it is being used."<sup>6</sup>

CEEB adopts a fairly narrow definition of SAT measurement objectives. SAT is intended only for

**Table 1.** Statistical Data Used in HAPLR\*

Statistic	Description	Type
Number of staff	Total paid FTE employees	Input
Materials expenditures	Expenditures on library collection	Input
Operating expenditures	Total operating expenditures (includes staff, collection, other non-capital expenditures)	Input
Number of volumes	Print materials	Input
Serial subscriptions	Print serial subscriptions	Input
Visits	Annual count of visits	Output
Reference transactions	Annual count of reference transactions	Output
Circulation transactions	Annual circulation count	Output

\*Fifteen HAPLR factors are created from just eight NCES statistical items. HAPLR also uses population and hours open to calculate rates (e.g., visits per hours open), but libraries are not rated on these two items.



**E**ven though libraries are urged to make statistical comparisons with their peers, the profession has no tried-and-true methods for judging adequacy or deficiency of any given statistical indicator.

involves methodological compromises. As another example, national ratings like HAPLR ignore imprecision in library statistical data because adjusting for noise in these data is typically impractical. Thus, aggregated library statistics remain inexact due to inconsistencies in data compilation, sampling error, and other reasons.

As indicated previously, the collective nature of national ratings makes them especially insensitive to local library uniqueness and variation. HAPLR attempts to address this variation by rating libraries within community population categories. Other ratings systems may use different grouping categories. However, these methods do not account for such important differences as community demographics, library mission, service responses, regional economic factors, funding disparities, and others.

Perhaps the most profound methodological compromise connected with national ratings, and with comparative library statistics in general, is what has been described by statisticians as *the creation of equivalences*. Standardized statistical definitions require that a variety of objects or events be combined into categories that emphasize similarities and ignore essential differences between the objects or events. In the case of library statistics, all volumes, subscriptions, dollars expended, in-person visits, website visits, personnel, programs, and so forth are presumed to be equivalent within each statistical category. Each is worth one "point" in the counting. This overarching assumption discounts differences in complexity, sophistication, value, quality, relevance, efficiency, effectiveness, and so on. We leave this and other methodological compromises for a future discussion in order to address another topic, the usefulness of single-score ratings.

### Single-Score National Ratings

Even if we could resolve the shortcomings and methodological compromises we noted, another dilemma remains. Hennen alluded to this in his characterization of national ratings of the "best" cities, hospitals, colleges, and graduate schools as "subjective and

open to infinite interpretation."<sup>11</sup> In reality, there are no objective guidelines for designing measures to gauge the excellence of cities, colleges, graduate schools, hospitals, libraries, and the like. Creators of national rating systems make arbitrary decisions about which factors to include or omit, and whether to give extra emphasis to any of the factors they select. Intentionally or not, each scoring decision favors certain locales or institutions over others.

When the *Places Rated Almanac* first appeared in 1985, AT&T Bell Laboratories statisticians analyzed the measurement scheme.<sup>12</sup> The statisticians examined weightings—a statistical technique for emphasizing some measurement dimensions over others—that the *Almanac* used. In an experiment, they found that by manipulating these weightings they could cause any one of 134 cities (of 329 total) to rate first in the rankings. A recent study described similar patterns in national rankings of regional economic conditions.<sup>13</sup> This study reports that thirty-four U.S. states are able to cite national rating systems that place them in "the top ten" in business climate and competitiveness.

While national rankings are arbitrary, they can also be somewhat imprecise. Year-to-year behavior of rankings can even be transitory. For graduate business school ratings published by *Business Week* and *U.S. News and World Report*, University of Michigan accounting professor Ilia Dichev observed that annual changes in ratings revert back to their prior ordering within a two-year period.<sup>14</sup> The same proved true for *U.S. News and World Report* college and university rankings.<sup>15</sup> Dichev suggests that this tendency is due to statistical noise, meaning that short-term changes in ratings are likely to be more of a fluke than an indication of actual changes in institutional quality.

Certainly, marketing and public relations are crucial for cities, universities, hospitals, libraries, and other institutions. National rankings can be valuable tools for institutional advocacy. By the same token, public institutions have a responsibility to increase operational efficiency, quality, and effectiveness. Due to their generic nature, national ratings do not

## A Primer on HAPLR Calculations

HAPLR calculations begin by assigning libraries to peer comparison groups based on community population size. Then, for each library, HAPLR combines eight library statistics (see table 1 on page 37) in different ways to produce fifteen separate measures, which Hennen calls "factors."<sup>1</sup> For instance, the single statistic *total annual library visits* is used in the measure *visits per capita* and also in the measure *visits per hours open*. Next, each library's fifteen measures are compared with those in its peer group. Example results from this step appear in the first row of table 2 for fictitious Cityburg Public Library. Because all fifteen measures are calculated in the same way, only four measures appear in table 2 to save space.

In table 2, note that Cityburg Public Library's *expenditures per capita* measure is \$11.50. For this value the library receives the rank of 147th place as shown in the second row of the table. In this manner, comparisons are made for measure #2, measure #3, and so forth so that a ranking is calculated for each of the fifteen measures.

HAPLR calculations then proceed using the fifteen rankings depicted in the second row of table 2. In the first row of the table, the actual statistical measures are discarded. Finally, the fifteen rankings are added together as shown here:

$$\text{CITYBURG PUBLIC LIBRARY HAPLR SCORE} = 147\text{th place} + 210\text{th place} + 94\text{th place} \dots + 87\text{th place} \\ = 650 \text{ points}$$

Actually, this formula is a simplification, although it accurately describes how HAPLR works:<sup>2</sup> Each library's fifteen measures are replaced with fifteen rankings and then the rankings are summed. As a result, final HAPLR scores are sums of rankings. The scores are still ordinal data because the fifteen rankings from which they are derived are ordinal data.

Ordinal data do not represent standard units of measure. Rather, ordinal data are irregular and elastic. One way to think of an ordinal scale is as a peculiar sort of yardstick. With this yardstick a given length—say ten inches—located at one end of the stick can be a different length than ten inches located at the other end! And ten inches in the middle of the yardstick can be a length that differs from the other two ten-inch lengths. Similarly, the difference between two scores from HAPLR's measurement scale—say, ten points between 420 and 430—is not necessarily equal to a corresponding difference elsewhere on that scale—say, points between 520 and 530 or between 660 and 670. Since ordinal scales have units of measure that are elastic, score differences that appear to be equal may not be equal at all.

### The Oddities of Ordinal Numbers

Why do ordinal numbers represent such irregular and elastic quantities? A familiarity with rankings and contests can help us solve this riddle. Rankings in contests do not directly indicate contestant perfor-

**Table 2.** HAPLR Calculations for Fictitious Cityburg Public Library\*

Cityburg Public Library	HAPLR Measure #1: Expenditures Per capita	HAPLR Measure #2: Percent budget to materials	HAPLR Measure #3: Materials expenditures per capita	HAPLR Measure #15: Circulation per visit
Input/output statistics	\$11.50	10 percent	\$1.09	2.7
Rankings	147th	210th	94th	87th

\*Input/output statistics are replaced by the library's ranking among peer libraries. Measures #4-#14 (not shown) are calculated in the same manner as the measures shown.

mances. That is, they do not represent actual timings in races or judging points awarded in dance competitions, for example. In a highly competitive contest, participants might all perform excellently, causing the top scores to fall quite close to each other. In a less competitive contest, top scores will be further apart and mediocre performances might well receive high rankings. Nevertheless, in these different contests top performers receive the same numeric rankings—first, second, third, and so on. Clearly, the numeric rankings have quite different meanings in different contests.

As ordinal numbers, rankings communicate order only—nothing more. Although it is tempting to view the numbers in rankings—first, second, third, fourth, and fifth—as regular numbers, the digits are merely placeholders that spell out a sequence. This sequence can just as easily be represented as Ath, Bth, Cth, Dth, and Eth. These two representations of a five-point ranking scale are identical in meaning, since points on ordinal scales are just labels.

Ordinal scaling is entirely different from everyday measurements. Because they are just labels, ordinal scaling points have no standard quantitative meaning. This is why HAPLR scores cannot indicate exact quantities that libraries possess of greatness, excellence, quality, or whatever library attributes HAPLR might actually measure. One library's advantage of, say, twenty HAPLR points over peer libraries could mean that the library surpasses other libraries by a large amount of excellence, by a moderate amount, by a small amount, or hardly at all. This uncertainty also applies to comparisons between libraries in different peer groups, and between a single library's year-to-year changes in HAPLR scores. It is even conceivable that one library's lower score in one year could represent more excellence than that library's higher score in a different year.

Due to the oddities of ordinal numbers, we can draw only these conclusions from HAPLR scores: In a single peer group and single year only, higher HAPLR scores exceed lower ones by an unspecified amount, and vice versa. Although the ratings appear to be precise numbers that can be compared confidently, their actual meanings are vague and fuzzy.

## What's the Point?

You might wonder why we belabor this issue of HAPLR scores as ordinal data. After all, no measurement system is perfect. Couldn't we simply accept the shortcomings of library measures like HAPLR? Despite its imperfections, isn't HAPLR the best measure of its kind currently available? In our view, such a lenient stance will ultimately work to the detriment of libraries. Instead, we suggest that libraries endeavor to promote higher standards of measurement soundness in order to preserve their institutional credibility. At a minimum, the profession ought to avoid showcasing measures that are likely to embarrass libraries later on. And we should strive to remedy measurement flaws that would otherwise be obvious to expert observers. As Elliot and his colleagues advise, library measures must hold up under expert critical review, particularly by those outside of our profession.<sup>3</sup> Further, libraries should always interpret assessment findings conservatively, tending toward understatement rather than overstatement. If the library profession upholds standards for rigorously developed assessment measures and justifiable interpretations of data, our "stewardship, service, and integrity" will be obvious to our constituents.<sup>4</sup> ■

## References and Notes

1. Thomas Hennen, "Go Ahead Name Them: America's Best Public Libraries," *American Libraries* 30, no. 1 (1999).
2. HAPLR calculations adjust the data so that numerically low rankings, like first and second place, earn more points than higher rankings, like eightieth and eight-first place. For clarity we have omitted this adjustment from our formula. With or without this adjustment, the numbers in the formula are ordinal rankings.
3. Donald S. Elliot et al., *Measuring Your Library's Value: How to Do a Cost-Benefit Analysis for Your Public Library* (Chicago: ALA Editions, 2007).
4. *Ibid.*, 36.



enhance local evaluation efforts, nor are they much help in refining programs and services. In fact, they sometimes undermine the practice of sound evaluation research. In response to *U.S. News & World Report* college and university rankings, some academic institutions choose to "game" the ratings by reporting exaggerated statistics. Defensive tactics like these end up obscuring organizational performance rather than illuminating it. Undoubtedly, this level of preoccupation with otherwise arbitrary rankings distracts organizations from the task of conducting well-designed local evaluations aimed at performance improvement.

## Telling Which Way the Wind Blows

Hennen remarks that "Numbers alone can no more tell you everything about a truly great library than the wind chill index can tell you about a truly cold day."<sup>16</sup> This disclaimer notwithstanding, national ratings such as HAPLR are perceived as saying *something* about each public library. The challenge for the library profession is to determine exactly what this something might be. Without scientific validation and sufficient measurement scope, public library national ratings based on aggregate statistics are unsatisfactory measures of quality, greatness, goodness, excellence, or value. Further, imprecision in the underlying data make ratings inexact, causing final scores to be much more approximate than they appear. Even if they were accurate, single-score ratings are not useful for meaningful performance, improvement, and accountability. Worse, the ratings can be interpreted in ways that misrepresent or trivialize organizational quality and excellence.

What, then, do library ratings say? Generally, our profession recognizes that ratings are rudimentary measures. The ratings may well signal library accomplishments, but the information they provide is preliminary and cursory. Library ratings indicate that libraries meet some of the prerequisites for quality, excellence, greatness, and value. Thus, they can suggest that libraries appear to be on the right track to performance results, if we accept the definitions of performance implied by the mix of the statistics utilized. By themselves, however, ratings cannot confirm these results.

Richer sources of information for public library management, marketing, and advocacy are soundly designed local evaluation studies, including customer surveys, performance measures, assessments of library value, quality management regimens, outcome studies, and others. For this reason, we pro-

pose that libraries continue to utilize data that are available from national surveys as a foundation for local assessment efforts, and add to these additional output measures that are currently not collected nationally.<sup>17</sup> Especially, the use of e-resource outputs (numbers of downloaded articles, webpage hits, database users, and so on) should be included. Data such as these are key to demonstrating how libraries add value to their communities. More importantly, libraries should strive to develop outcome measures—both quantitative and qualitative—in their assessments of how they fall short, meet, or exceed organizational targets and goals. National ratings can play a part in library evaluation as long as we acknowledge their strengths and limitations. Beyond their use in library advocacy efforts, ratings should inspire libraries to identify local evaluation questions needing further exploration.

Utilizing local data to demonstrate library accomplishments is worthwhile because it focuses on locally derived value of library services. The library's ability to return value to customers is one rating system that libraries must be well attuned to. Customer opinions and votes are the only scores that take away or add to the resources libraries require to operate at a high level and to strive for greatness. ■

*Editor's note: The opinions expressed in this article are those of the authors and do not reflect the view of the National Oceanic and Atmospheric Administration or other parts of the U.S. government.*

## References and Notes

1. Hennen decided to forgo publishing 2007 HAPLR ratings.
2. Thomas Hennen, "Go Ahead Name Them: America's Best Public Libraries," *American Libraries* 30, no. 1 (1999): 76; —, "Great American Public Libraries: HAPLR Ratings 2000," *American Libraries* 31, no. 10 (2000): 51.
3. Glen Holt, "What Makes a Library Great?" *Public Library Quarterly* 24, no. 2 (2005): 83–89; Thomas Childers and Nancy A. Van House, "The Grail of Goodness: The Effective Public Library," *Library Journal* 114, no. 16 (1989): 44–49.
4. Ray Lyons, "Unsettling Scores: An Evaluation of Hennen's Annual Public Library Ratings," *Public Library Quarterly* 26, no. 3/4 (2007): 49–100.
5. Since its scores are comprised of nine output and six input measures, HAPLR appears to stress output measures more. However, based on statistical weightings built into the calculations,

- the HAPLR scores are made up of 50 percent input and 50 percent output measures. See Lyons, "Unsettling Scores," 58–60.
6. Edward G. Carmines and Richard A. Zeller, *Reliability and Validity Assessment* (Beverly Hills, Calif.: Sage, 1979): 17.
  7. Rebecca Zwick, "College Admission Testing," National Association for College Admission Counseling, 2007, [www.nacacnet.org/NR/rdonlyres/26606B59-5725-4A4B-9203-8D12D805B7C8/0/FinalStandardizedTestingWhitePaper.pdf](http://www.nacacnet.org/NR/rdonlyres/26606B59-5725-4A4B-9203-8D12D805B7C8/0/FinalStandardizedTestingWhitePaper.pdf) (accessed Apr. 1, 2008).
  8. "Standardized Testing Bibliography," National Association for College Admission Counseling, 2006, [www.nacacnet.org/NR/rdonlyres/B16EAF26-3362-405F-A2B0-8CCD9A2A3EDD/0/Testing\\_Bibl.pdf](http://www.nacacnet.org/NR/rdonlyres/B16EAF26-3362-405F-A2B0-8CCD9A2A3EDD/0/Testing_Bibl.pdf) (accessed Apr. 1, 2008).
  9. For details on HAPLR calculations see Ray Lyons, "Unsettling Scores," 52–58.
  10. Ellen Altman, "Reflections on Performance Measures Fifteen Years Later," in *Library Performance, Accountability, and Responsiveness: Essays in Honor of Ernest R. DeProspero*, ed. Charles C. Curran and F. William Summers (Norwood, N.J.: Ablex, 1990): 9–16.
  11. Hennen, "Go Ahead Name Them," 72.
  12. Richard A. Becker et al., "Analysis of Data from the Places Rated Almanac," *The American Statistician* 41, no. 3 (1987): 169–86.
  13. Peter S. Fisher, *Grading Places: What Do the Business Climate Ranks Really Tell Us?* (Washington, D.C.: Economic Policy Institute, 2005), [www.epinet.org/books/grading\\_places/grading\\_places\\_full\\_text.pdf](http://www.epinet.org/books/grading_places/grading_places_full_text.pdf) (accessed Apr. 1, 2008).
  14. Iliia D. Dichev, "How Good Are Business School Rankings?" *The Journal of Business* 72, no. 2 (1999): 201–13.
  15. ———, "News or Noise? Estimating the Noise in U.S. News University Rankings," *Research in Higher Education* 42, no. 3 (2001): 237–66.
  16. Thomas Hennen, "Great American Public Libraries: HAPLR Ratings Round Two," *American Libraries* 30, no. 8 (1999): 68.
  17. Input/output statistics from the Public Library Statistics Cooperative (formerly the Federal-State Cooperative System) may be accessed via the Institute of Museum and Library Services (<http://harvester.census.gov/imls/index.asp>), the National Center for Educational Statistics ([www.nces.ed.gov/surveys/libraries/public.asp](http://www.nces.ed.gov/surveys/libraries/public.asp)), and the Public Library Geographic Database of the Florida State University College of Information Science ([www.geolib.org/PLGDB.cfm](http://www.geolib.org/PLGDB.cfm)) (all accessed Oct. 27, 2008).

### *Flight of the Conchords* films at Brooklyn Public Library

Recently, the cast of HBO's *Flight of the Conchords* were all at Brooklyn Public Library's Williamsburgh Branch, filming for their upcoming season, which airs in January. In the scene, the duo performs—not surprisingly—in a library, to the chagrin of regular patrons. After this portion of the shoot, they headed to Sunset Park to film more of the episode in an area under the Brooklyn-Queens Expressway, and brought their own hookers (well, extras dressed as hookers, anyway).

Tracey Mantrone, manager of the Williamsburgh Branch, had this to say:

Although I don't watch much television, my sister is a big fan of *Flight of the Conchords*. I was therefore happy that not only was her favorite show filming in my library, but I would be able to watch. Little did I know that I would find myself not only dismantling most of my computers, but also helping to shelve books. The film crew, while struck by the beauty of the building, decided that it did not look enough like a library. So they replaced the computers with temporary book shelves, which we stuffed full of donations to create an old-time atmosphere. Then the extras, dressed in frumpy cardigan outfits and clunky shoes and wearing owl-like glasses, began to arrive. By the time filming began, the entire building had developed a carnival-like atmosphere as a bemused public complimented us on the extra shelving and waded among the film crew to check out their DVDs for the weekend.

Then the filming began. For the next hour and a half, we all suffered through a bizarrely catchy song sung over and over again by the two lead actors while the extras did their best to look like traditionally frumpy librarians. . . . I'm looking forward to watching the upcoming episode, although I still can't get that song out of my head. ☐





## COPYRIGHT INFORMATION

TITLE: Honorable Mention: What Public Library National  
Ratings Say

SOURCE: Public Libr 47 no6 N/D 2008

The magazine publisher is the copyright holder of this article and it is reproduced with permission. Further reproduction of this article in violation of the copyright is prohibited. To contact the publisher:  
<http://www.ala.org>